

Informatique et cartographie

Serge Lhomme

Maître de conférences en géographie

<http://sergelhomme.fr/>

serge.lhomme@u-pec.fr

15 décembre 2018

- 1 Introduction : Les fondements de la cartographie
- 2 Dessiner une carte vectorielle : La DAO
- 3 Produire des cartes thématiques quantitatives : Les outils spécialisés
- 4 Analyser de l'information géographique : Les tableurs

- 1 Introduction : Les fondements de la cartographie
- 2 Dessiner une carte vectorielle : La DAO
- 3 Produire des cartes thématiques quantitatives : Les outils spécialisés
- 4 Analyser de l'information géographique : Les tableurs

Généralités sur l'information géographique

Définition

Définition

L'information géographique est la représentation d'un objet ou d'un phénomène localisé dans l'espace à un moment donné.

Il s'agit d'un type d'information très répandu, décrivant des objets, phénomènes, êtres vivants ou sociétés, dès lors qu'ils sont reliés à un territoire.

Généralités sur l'information géographique

Les deux composantes de l'information géographique

Les deux principales composantes de l'information géographique sont :

- Le niveau sémantique qui correspond à l'information relative à un objet. L'ensemble des attributs de l'objet forme ses attributs (comme par exemple : le numéro d'une parcelle cadastrale, le nom d'une route, le nombre d'habitants d'une commune. . .).
- Le niveau géométrique qui correspond à la forme et à la localisation de l'objet sur la surface terrestre. Un système de coordonnées peut être valable sur tout ou partie de la surface terrestre. On peut aussi définir un système de « coordonnées relatives » par rapport à un point d'origine quelconque, comme c'est souvent le cas pour les relevés topographiques.

Généralités sur l'information géographique

Les formes de représentation : distinctes mais complémentaires

Pour représenter l'information géographique, il existe trois formes :

- L'information géographique peut être représentée sur une image où l'on peut voir une multitude d'objets (comme par exemple une photo aérienne ou une image satellite) sans connaître directement leurs attributs (on ne voit pas le nom de la route, le nombre d'habitants des communes...).
- L'information géographique se prête particulièrement bien à la représentation sur une carte où l'on situe les objets et les phénomènes dans un repère général et homogène.
- L'information géographique peut être représentée par un texte ou un fichier de données littérales où elle est représentée par des données numériques et par une adresse (comme par exemple les fichiers des abonnés à l'ADSL qui contiennent nom, prénom, numéro de téléphone, adresse postale. . .).

Généralités sur l'information géographique

Les formes de représentation : distinctes mais complémentaires



Image
Géométrie
sans sémantique



Carte
Géométrie
et sémantique

Nom de lieu	hab.
la Croix Bertho	32
la Guette	115
Pommeret	949
les Prés	47
Goulven	78
Bellevue	8
Carnopern	15
Sainte-Anne	56

Texte
Sémantique
sans géométrie

Généralités sur l'information géographique

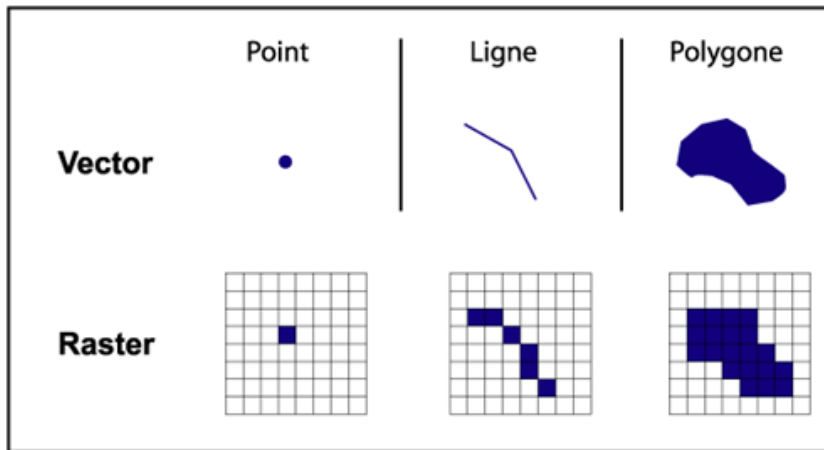
L'information géographique d'un point de vue numérique

D'un point de vue numérique, il existe deux modes de représentation :

- Le mode maillé (ou raster en anglais), où la surface de la carte ou de l'image est décrite selon un balayage ligne par ligne. Chaque ligne est composée de pixels. C'est mathématiquement, ni plus ni moins, qu'une matrice (dans le langage usuel un tableau de chiffres).
- Le deuxième mode de représentation est le mode vecteur, où chaque objet représenté sur la carte est décrit par un ou des points successifs composant sa forme. Chaque point est localisé par ses coordonnées rectangulaires. La forme des objets est exprimée par le biais de trois principales « primitives » géométriques : objets ponctuels, linéaires et surfaciques.

Généralités sur l'information géographique

L'information géographique d'un point de vue numérique



Les fondements de la cartographie

La subjectivité

On appelle carte toute représentation graphique partielle ou complète dans le plan d'un objet plus complexe. Dans les domaines de la géographie et de l'aménagement, cet objet est un territoire représenté selon une "vue de dessus".

Cette représentation graphique est établie par un auteur, à un moment donné, sur un espace donné. Ce caractère interprétatif (voire subjectif) de la conception cartographique est largement accepté par les géographes.

D'autres publics imaginent parfois que l'image du territoire représentée sur une carte est véridique et même objective.

Naturellement, aucun cartographe digne de ce nom ne cherche à tromper ses lecteurs, mais force est de constater que le travail du cartographe résulte de multiples choix d'application, de conventions, de plaisirs esthétiques qui ne sont pas toujours explicités, ni justifiés.

Les fondements de la cartographie

L'art cartographique

La cartographie peut être définie comme l'« ensemble des études et des opérations scientifiques, artistiques et techniques intervenant à partir des résultats d'observations directes ou de l'exploitation d'une documentation, en vue de l'élaboration de cartes et autres modes d'expression, ainsi que dans leur utilisation ».

A l'instar des mathématiques, la cartographie tend à être un langage universel. Ainsi quelle que soit la langue utilisée, une carte est théoriquement compréhensible par tout le monde, même si la légende est parfois nécessaire pour saisir des points de détails.

Pour cela, ce langage doit respecter les règles de lisibilité, de clarté, d'intelligibilité et d'enchaînement logique, inhérentes à tout langage humain.

Les fondements de la cartographie

L'objectif du cartographe

Outil de communication par l'image, la carte doit être perçue avec un minimum de biais, dans la mesure où le concepteur a su prendre en compte les lois de la perception visuelle, du pouvoir intégrateur et séparateur de l'œil, des contrastes de couleurs, et des règles typographiques concernant les écritures.

Entre la subjectivité de l'objet qu'il réalise et la possibilité d'en faire un objet d'art, le cartographe ne doit jamais oublier que sa première mission est de se faire comprendre pour apporter l'information souhaitée.

Comme il est toujours difficile de se faire comprendre, il ne faut jamais oublier de légender sa carte.

Les fondements de la cartographie

Qu'est-ce que dessiner carte ?

Dessiner une carte, c'est tracer la forme d'éléments cartographiques (des objets géographiques) en fonction d'une échelle choisie de telle manière que le document final puisse être tenu en main.

Ces éléments cartographiques sont des représentations d'objets géographiques matériels (routes, maisons...) ou immatériels (limites communales, frontières nationales), visibles (aéroports) ou invisibles (réseaux enterrés), immobiles ou mobiles (camions).

La représentation des éléments cartographiques fait appel à trois types de tracés (ou primitives graphiques, ou encore structures visuelles) : le point, le segment (la ligne ou la polyligne), le périmètre fermé et sa surface associée (un polygone).

Les fondements de la cartographie

Primitives, structures visuelles

Un objet géographique matériel prend la forme de telle ou telle autre primitive graphique en fonction de l'échelle de représentation : à l'échelle du 1 : 1 000 000 une ville est représentée par un point, alors qu'à l'échelle du 1 : 100 000, elle occupe un périmètre plus ou moins nettement délimité.

Points



Lignes



Surfaces



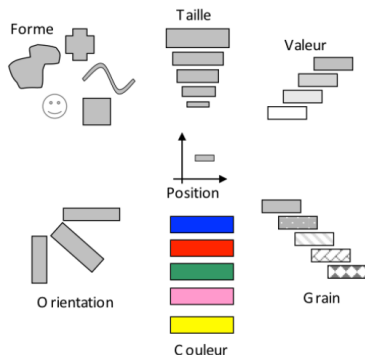
Volumes



Les fondements de la cartographie

Coder les informations

Pour coder des informations, il est possible de faire varier certaines propriétés graphiques de ces structures visuelles (par exemple la forme). Les variations possibles sur les structures visuelles sont regroupées par type et sont nommées « variables visuelles ».



Les fondements de la cartographie

La carte est un ensemble de calques

Du point de vue de l'infographie, une carte thématique peut être considérée comme l'empilement d'un ensemble de calques.

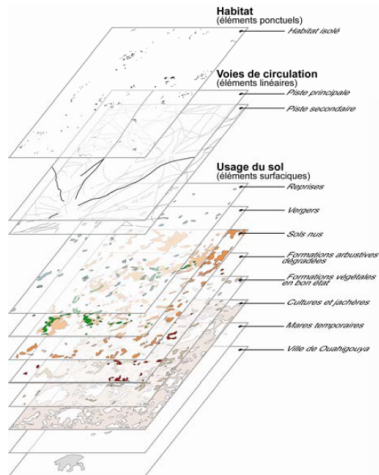
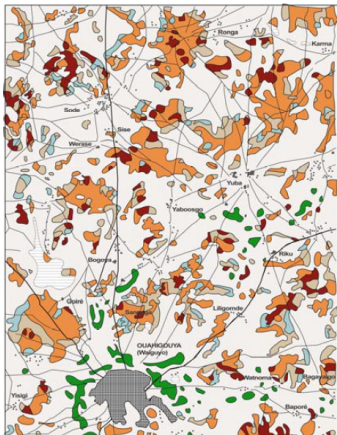
Chacun de ces calques contient un type d'éléments cartographiques et un seul.

Contrairement au lecteur, le cartographe doit voir la carte comme un ensemble de calques et non comme un ensemble d'objets.

Le parallèle entre la décomposition de la carte en différents calques et la lecture de la légende est alors souvent évident. En effet, une structure visuelle combinée à des variables visuelles correspond à un type d'objets souvent souvent associé à un calque.

Les fondements de la cartographie

La carte est un ensemble de calques




Les fondements de la cartographie

La carte parfaite n'existe pas

Cultivé


 Cultures et jachères

Non Cultivé

 Formation végétale en bon état

 Formation végétale dégradée

 Sol nu

 Reprises le long des ouvrages

 Verger

 Mare

 Périmètre urbain de OAHIGOUYA

 Habitat isolé

 Piste principale


 Piste secondaire

 Limite de terroir

Cultivé

 Cultures et jachères

Non Cultivé

 Formation végétale en bon état

 Formation végétale dégradée

 Sol nu

 Reprises le long des ouvrages


 Verger

 Mare

 Périmètre urbain de OAHIGOUYA

 Habitat isolé

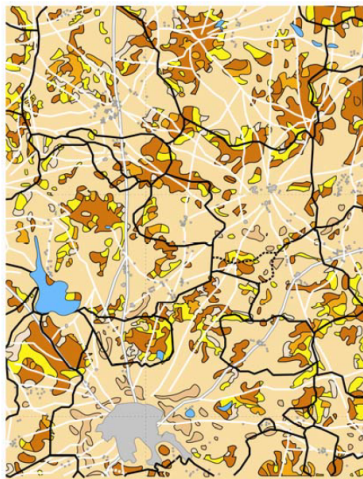
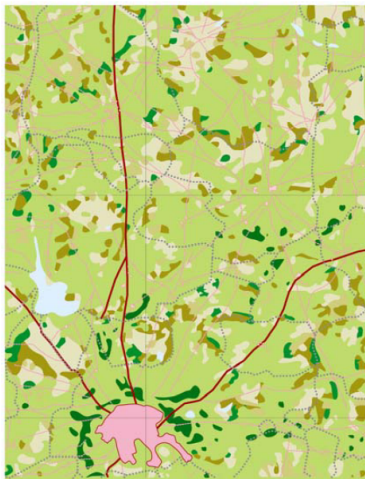
 Piste principale

 Piste secondaire

 Limite de terroir

Les fondements de la cartographie

La carte parfaite n'existe pas



Les fondements de la cartographie

Faire une carte ne se limite pas à produire une simple image artistique d'un espace. La carte doit transmettre un message. Pour cela, il faut :

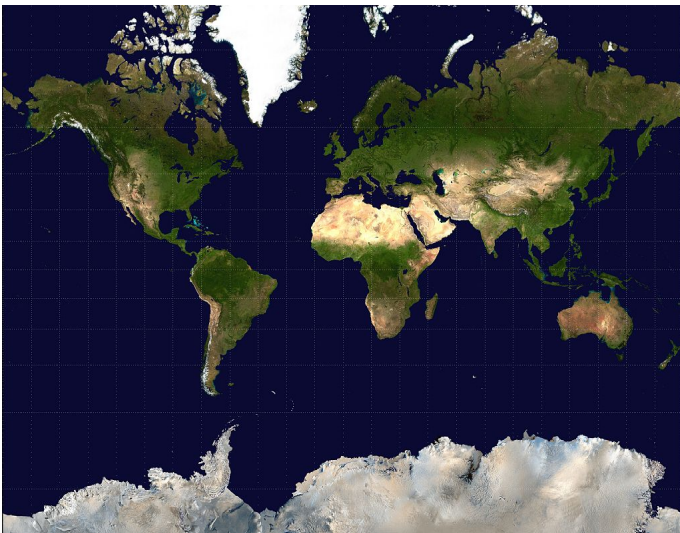
- Identifier l'objectif de la carte ;
- Identifier la cible ;
- Identifier l'information à cartographier (collecte et traitement) ;
- Adapter le fond de carte (projection, généralisation) ;
- Choisir le langage cartographique.

Il ne faut pas oublier :

- D'indiquer le nord ;
- De préciser l'échelle ;
- De mettre une légende ;
- De préciser les sources ;
- De mettre un titre.

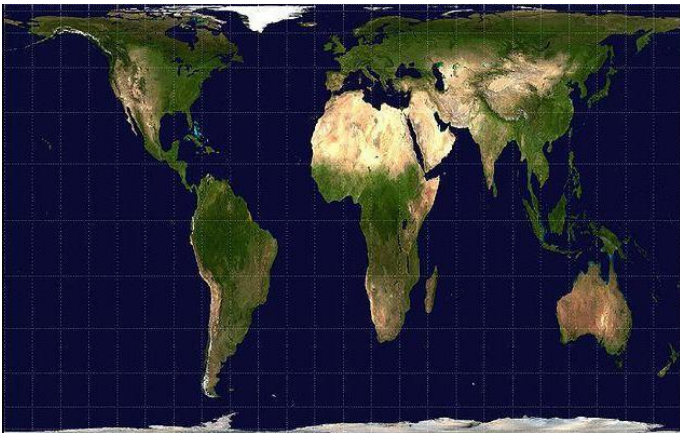
Toutes les cartes sont fausses

La projection Mercator (1569)



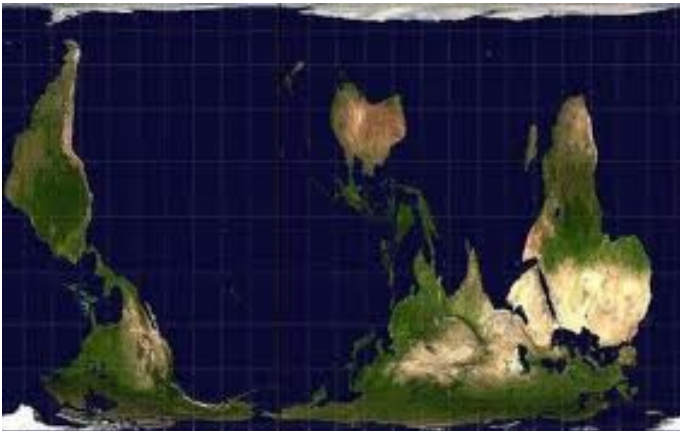
Toutes les cartes sont fausses

La projection de Peters (1967)



Toutes les cartes sont fausses

Mercator Australie



Toutes les cartes sont fausses

Stéréographique : conforme



- 1 Introduction : Les fondements de la cartographie
- 2 Dessiner une carte vectorielle : La DAO**
- 3 Produire des cartes thématiques quantitatives : Les outils spécialisés
- 4 Analyser de l'information géographique : Les tableurs

Maitriser un outil de DAO

Qu'est-ce que la DAO ?

Le dessin assisté par ordinateur (DAO) est une discipline permettant de produire des dessins techniques avec un logiciel informatique. De ce fait, en DAO, la souris et le clavier remplacent le crayon et les autres instruments du dessinateur.

Les dessins produits sont le plus souvent réalisés en mode vectoriel (traits cohérents).

L'intérêt de la DAO, est d'abord celui de l'informatique, c'est-à-dire essentiellement un apport de praticabilité dans la gestion des documents, facilitant l'édition de modifications, l'archivage, la reproduction, le transfert de données, etc.

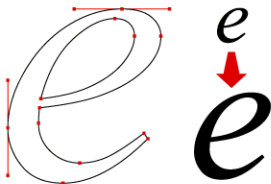
Deux exemples : Adobe Illustrator et Inkscape.

Maitriser un outil de DAO

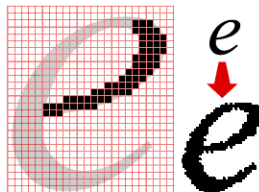
Pourquoi dessiner en vectoriel ?

Le principal avantage des images vectorielles est qu'elles sont indépendantes de la résolution, c'est-à-dire qu'elles gagnent en qualité si on les agrandit pour mieux les observer. Il n'y a pas de pixelisation, car la qualité de l'image s'adapte à la résolution de l'écran.

VECTOR GRAPHICS



BITMAPMED (RASTER) GRAPHICS



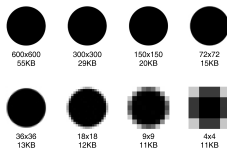
Maitriser un outil de DAO

Pourquoi dessiner en vectoriel ?

Par nature, un dessin vectoriel est dessiné à nouveau à chaque visualisation, ce qui engendre des calculs sur la machine. Ce qui peut engendrer des temps d'affichage important pour des dessins volumineux.

Il est possible de modifier chaque ligne très rapidement et d'un seul coup. Chaque élément est ainsi indépendant et une erreur peut être vite corrigée. Même une conception brouillonne peut donner un résultat satisfaisant.

De surcroit, cette représentation implique un nombre finalement limité d'objets graphiques, les fichiers graphiques sont très légers.



Maitriser un outil de DAO

Comment dessiner en vectoriel ?

Pour lire des formats vectoriels, il faut des logiciels spécifiques pour effectuer les calculs. Heureusement, un format tend à devenir de plus en plus répandu : le SVG. De nombreux logiciels permettent ainsi de le lire, comme les navigateurs internet.

Comme le SVG est un format ouvert, si on maitrise son langage, il est possible de dessiner en écrivant !!!

Un exemple

```
<svg xmlns="http://www.w3.org/2000/svg" version="1.1" width="300"
height="200">
<rect width="100" height="80" x="0" y="70" fill="green" />
</svg>
```

Application à la cartographie

L'objectif est ici de réfléchir sur l'accessibilité spatiale de l'université de Créteil. En effet, cette université en pleine mutation est initialement une université de "proximité", créée pour faciliter l'accès à l'université à un public d'étudiants n'y ayant pas initialement accès.

Dans ce cadre, il s'agit premièrement de produire une carte de situation de l'université. Cette carte à destination de gestionnaires a pour objectif de présenter l'accessibilité spatiale du site. Cette carte pourra être ensuite insérée au sein d'un document synthétisant une réflexion sur cette problématique (un poster).

- Importez sous Illustrator le fond de carte disponible.
- Reproduisez en partie cette carte et améliorez celle-ci afin d'appréhender au mieux la problématique.

- 1 Introduction : Les fondements de la cartographie
- 2 Dessiner une carte vectorielle : La DAO
- 3 Produire des cartes thématiques quantitatives : Les outils spécialisés**
- 4 Analyser de l'information géographique : Les tableurs

Produire des cartes thématiques quantitatives

Enrichir l'information géographique

Réaliser une carte, ce n'est pas toujours essayer de représenter au mieux les territoires qui nous entourent, de les dessiner, comme dans les cartes topographiques et les cartes d'inventaire.

Réaliser une carte, c'est parfois mettre en évidence des caractéristiques que l'on ne voit pas forcément.

Pour cela, il convient de ne pas se contenter de représenter l'information géographique, il convient plutôt de l'analyser et de la rendre intelligible.

Il existe deux grandes méthodes d'analyse des données géographiques : l'analyse quantitative et l'analyse qualitative.

On va se concentrer ici sur l'analyse quantitative. En effet, c'est l'analyse la plus simple. Pour cela, on s'appuiera sur la sémantique de l'information géographique : le tableau d'information géographique.

Produire des cartes thématiques quantitatives

Le tableau d'information géographique : la sémantique de l'information géographique

Variables Caractères

	AGRI	ARTI	CADRE	PRO INT	EMPLOYE	OUVRIER	RETRAITE	AUTRES
0	4067	16745	36426	70663	77349	78998	30417	29910
1	5201	11414	18629	48889	71578	78906	32063	42108
2	6159	9396	11842	31102	46196	42101	24218	21344
3	2027	6387	6951	16140	20885	15560	10687	10278
4	2040	5286	5883	15687	19707	12415	8865	7222
5	1918	39514	73884	117652	160574	87227	54801	68990
6	4132	10758	12461	31859	39088	37199	21660	19063
7	3362	6520	9771	25487	36326	42618	16464	23226
8	2348	4888	5537	14442	20788	16635	10379	9741
9	4786	7198	12499	28889	40121	43224	18616	18373
10	5474	12045	13486	31504	47297	34501	25584	25482
11	11100	10116	9937	25717	33970	29390	18550	13039
12	4508	52145	135584	225526	268192	168920	100790	155952
13	6081	18540	36155	72740	94384	82643	41776	34559
14	7431	4976	4547	12517	19508	17183	10386	7786
15	5756	10236	14308	32982	46169	46763	24653	20416
16	9061	21025	24906	57199	84800	65577	46140	35151
17	4025	8183	12967	29897	42604	39209	22057	18361
18	4930	7553	9273	23538	33050	28048	16824	12970
19	5352	13503	32185	62097	71854	61993	30098	23303
20	12517	17359	24766	54660	70213	68805	43023	29034
21	5259	3696	3669	9387	16039	12764	9811	6780
22	7694	15460	14125	34308	54103	47943	31651	24558
23	4123	12196	29138	57994	65018	77588	28449	27692
24	6106	14998	23027	52929	59618	57491	28522	30398
25	4064	14930	28312	62716	74944	85833	34937	33452
26	3848	10110	22675	47678	58632	57416	25285	22370
27	11054	23043	46057	94797	116573	101817	59116	46699

Entités
Individus
Objets
Unités

Valeur
Modalité

Produire des cartes thématiques quantitatives

Variables et valeurs

Les variables d'un tableau d'information géographique peuvent être quantitatives ou qualitatives.

Une variable qualitative est composée de valeurs qui ne sont pas des quantités (ou des rapports de quantités). Les valeurs prises par cette variable ne peuvent pas faire l'objet d'opérations mathématiques simple comme une addition (en tout cas ça n'a pas de sens). Ces valeurs peuvent être simplement nominales (lettres, mots, chiffres) ou ordinales c'est-à-dire qu'on peut les ordonner par ordre croissant (les classements : 1er, 2ème, 3ème, ... ; les jugements : bons, moyens, mauvais...).

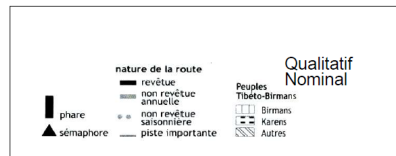
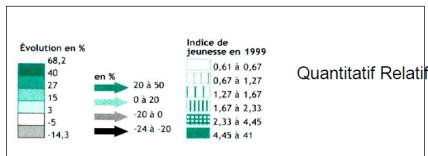
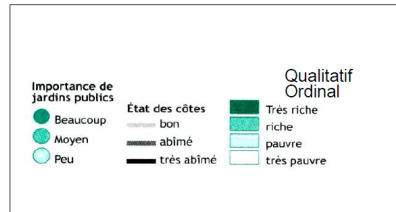
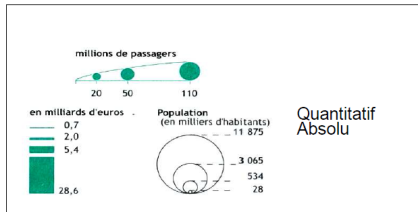
On va dans ce cours plutôt se focaliser sur des variables quantitatives. Les valeurs sont des quantités (des nombres d'individus). Toutes les opérations peuvent être opérés sur les valeurs d'une variable.

Produire des cartes thématiques quantitatives

Résumé

En cartographie, on différencie notamment les variables (quantitatives) de stock et celle de taux.

On distinguera aussi les variables de flux (tableau d'échange).



TP

Produire des cartes thématiques quantitatives

- ① Dans PhilCarto représentez la variable cadre. Quelle variable visuelle a été choisie pour représenter cette variable ? Est-ce juste ? Quelles sont les qualités et les défauts de cette carte ?
- ② Enregistrez cette carte au format ai.
- ③ Dans PhilCarto représentez la variable Tx_Cadre et Tx_Profession intermédiaire. Quelle variable visuelle a été choisie pour représenter cette variable ? Est-ce juste ? Quelles sont les qualités et les défauts de cette carte ?
- ④ Comparez les résultats obtenus avec 4 discrétisations différentes comportant toutes 5 classes. Enregistrez une synthèse au format EMF. Décrivez les principes de chaque discrétisation.

- 1 Introduction : Les fondements de la cartographie
- 2 Dessiner une carte vectorielle : La DAO
- 3 Produire des cartes thématiques quantitatives : Les outils spécialisés
- 4 Analyser de l'information géographique : Les tableurs

Discrétisation

Simplifier l'information pour faciliter le message à faire passer

La discrétisation consiste à découper des données en classes homogènes.

C'est une opération de simplification.

Pour réaliser une discrétisation, il faut déterminer le nombre de classes et les bornes des classes.

Pour réaliser une bonne discrétisation, il faut justifier à la fois le nombre de classes et les bornes de ces classes.

Intuitivement, un bon découpage correspond à des classes homogènes et séparées, ce qui correspond respectivement aux notions statistiques de faible variance intraclasse et de forte variance interclasse.

Et oui, pour réaliser une simple carte représentant une seule et pauvre variable, il faut des connaissances statistiques de base solide !

Discrétisation

Les résumés statistiques classiques

La moyenne :

$$\bar{X} = \frac{1}{N} \times \sum_{i=1}^N X_i$$

L'écart-type :

$$\sigma_D = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2}$$

Les caractéristiques de position fondées sur l'ordonnancement (les modes) :

Minimum ; Maximum ; Médiane ; Quantiles ; Déciles...

TP

Résumé statistique : application

Vous disposez d'un fichier Excel permettant d'obtenir la répartition de différentes catégories socioprofessionnelles au sein des départements de France métropolitaine.

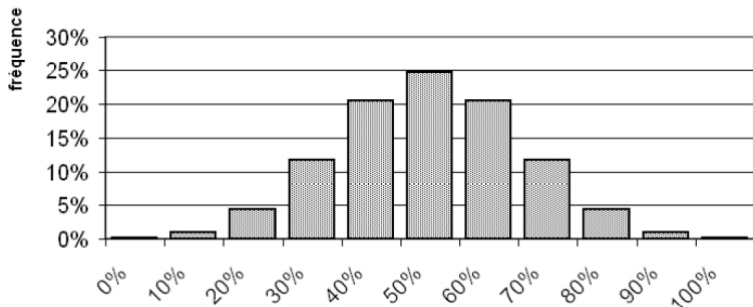
- 1 Ouvrez le fichier, puis calculez la somme, la moyenne et la médiane de chaque CSP.
- 2 Déterminez la variance, puis l'écart-type pour la variable Ouvrier.
- 3 Figez la première ligne afin de la visualiser de manière permanente.
- 4 Créez un tableau. Quel changement s'est opéré au niveau des formules si vous calculez par exemple le minimum ou le maximum de chacune des variables ?

Discrétisation

L'arme absolue : la distribution statistique

La distribution statistique (distribution des fréquences) est un tableau qui associe des classes de valeurs obtenues lors d'une expérience à leurs fréquences d'apparition. Ce tableau de valeurs est modélisé en théorie des probabilités par une loi de probabilité. Il peut notamment se représenter sous la forme d'un histogramme.

Distribution des réussites (10 questions)



Discrétisation

Les règles

Le nombre de classes est en théorie proportionnel au nombre d'individus observés. Ainsi il peut paraître excessif de créer 8 classes pour seulement 15 pays. Il existe un indice permettant de connaître le nombre idéal de classes.

$$N(cl) = 1 + 3,3 \log(N) \text{ Huntsberger}$$

L'objectif est bien souvent de conserver la forme de la distribution et donc de conserver au mieux la structure interne : Jenks ou Seuils naturels.

Pour illustrer la dispersion des variables étudiées, il faut tout simplement choisir une classification par amplitude égale (qui permet en fait d'obtenir une carte de la distribution statistique).

Pour comparer la position de certains lieux en fonction de différentes variables (sur plusieurs cartes), il faudra utiliser des méthodes faisant référence à des paramètres statistiques (moyenne, médiane, EcT...).

TP

Distribution et discrétisation

- 1 Dans Philcarto, on a testé différentes représentations obtenues par différentes discrétisations des variables suivantes : Tx_Cadre, Tx_Profession intermédiaire. La question laissée en suspens était la suivante : quelle est la meilleure représentation pour chaque variable ?
- 2 Réalisez des histogrammes de distribution statistique pour ces deux variables. Testez pour cela plusieurs amplitudes de classe. Quelles sont les formes de ces deux distributions statistiques ?
- 3 A partir de l'analyse des distributions obtenues, choisissez la meilleure méthode de discrétisation, puis choisissez le bon nombre de classes afin d'obtenir la meilleure représentation possible pour chaque variable.
- 4 Finalisez cette carte à l'aide d'Illustrator.

Coefficient de spécialisation

Poids des entités géographiques et des variables dans un tableau de contingence

Effectifs observés (N_{ij})									
1963	ALIM	TEXT	BOIS	EDIT	CHIM	CONS	META	EQUIP	Total
BULGARIE	130	128	39	14	29	47	21	151	559
HONGRIE	144	241	53	28	77	61	91	423	1118
POLOGNE	380	612	164	84	222	199	147	881	2689
R.D.A.	206	451	119	118	308	142	109	1056	2509
ROUMANIE	136	305	244	41	76	114	106	366	1388
TCHECO.	185	412	130	63	139	151	177	883	2140
YOUGOSL.	126	223	132	58	76	78	69	307	1069
Total	1307	2372	881	406	927	792	720	4067	11472

Coefficient de spécialisation

Deux profils possibles : le profil en ligne

Profils en ligne (Nij/Ni.)									
1963	ALIM	TEXT	BOIS	EDIT	CHIM	CONS	META	EQUIP	Total
BULGARIE	23%	23%	7%	3%	5%	8%	4%	27%	100%
HONGRIE	13%	22%	5%	3%	7%	5%	8%	38%	100%
POLOGNE	14%	23%	6%	3%	8%	7%	5%	33%	100%
R.D.A.	8%	18%	5%	5%	12%	6%	4%	42%	100%
ROUMANIE	10%	22%	18%	3%	5%	8%	8%	26%	100%
TCHECO.	9%	19%	6%	3%	6%	7%	8%	41%	100%
YOUGOSL.	12%	21%	12%	5%	7%	7%	6%	29%	100%
Total	11%	21%	8%	4%	8%	7%	6%	35%	100%

Coefficient de spécialisation

Deux profils possibles : le profil en colonne

Profils en colonne (Nij/N.j)									
1963	ALIM	TEXT	BOIS	EDIT	CHIM	CONS	META	EQUIP	Total
BULGARIE	10%	5%	4%	3%	3%	6%	3%	4%	5%
HONGRIE	11%	10%	6%	7%	8%	8%	13%	10%	10%
POLOGNE	29%	26%	19%	21%	24%	25%	20%	22%	23%
R.D.A.	16%	19%	14%	29%	33%	18%	15%	26%	22%
ROUMANIE	10%	13%	28%	10%	8%	14%	15%	9%	12%
TCHECO.	14%	17%	15%	16%	15%	19%	25%	22%	19%
YOUGOSL.	10%	9%	15%	14%	8%	10%	10%	8%	9%
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%

Coefficient de spécialisation

Explication

Le quotient de localisation est un indicateur de « concentration », de spécialisation.

Il donne une mesure de l'importance relative d'une modalité (ou catégorie) dans une unité spatiale, comparée à son poids dans les autres unités spatiales.

Le quotient de localisation est un outil d'analyse spatiale, car il permet de caractériser le degré de concentration d'une sous-population dans une unité spatiale en le comparant à toutes les autres unités spatiales d'un même ensemble territorial.

Il permet de mener cette comparaison en faisant abstraction des inégalités de poids entre les unités spatiales et les différentes catégories.

Coefficient de spécialisation

Présentation

	C1	C2	C3	Total
B1	x11	x12	x13	x1●
B2	x21	x22	x23	x2●
B3	x31	x32	x33	x3●
Total	x●1	x●2	x●3	x●●

$$Q(x[ij]) = (x[ij]/x[.j]) / (x[i.]/x[.]) = (x[ij] \times x[.]) / (x[i.] \times x[.j])$$

Coefficient de spécialisation

Exemple

Branches Zone	B1	B2	B3	Total
Z1	48	325	287	660
Z2	27	185	148	360
Z3	45	90	45	180
Total	120	600	480	1200

Coefficient de spécialisation

Exemple

Branche Zone	B1	B2	B3	Total
Z1	48	325	287	660
Z2	27	185	148	360
Z3	45	90	45	180
Total	120	600	480	1200

$$Q(X[z2b1]) = \frac{(27/120)}{(360/1200)} = \frac{0,225}{0,300} = 0,75$$

Coefficient de spécialisation

Exemple

Branche Zone	B1	B2	B3
Z1	0,727	0,985	1,087
Z2	0,750	1,028	1,028
Z3	2,500	1,000	0,625

Coefficient de spécialisation

Et quand le tableau n'est pas un tableau de contingence ? La standardisation

Name	DEM1	DEM2	DEM3	ECO1	ECO2	ENV1	ENV2
France	107	303	12	26200	19100	6.6	0.35
Italy	192	529	9	20100	18000	7.6	0.42
Spain	78	205	9	14500	14100	6.4	0.45
Algeria	13	385	30	1500	3000	3.3	1.12
Libya	3	238	28	2000	4800	8.8	1.83
Morocco	63	294	23	1300	3600	1.1	0.29
Tunisia	58	194	22	2100	5300	1.8	0.34
West. Medit.	38	310	15	14800	13000	5.5	0.42
Definition of variables							
DEM1	Gross population density in inh/km2 (POP/SUP)						
DEM2	Net population density in inh/km2 (POP/AGR)						
DEM3	Birth rate (BIR/POP)						
ECO1	GNP in \$ per inhabitant (GNP/POP)						
ECO2	GDP in p.p.a per inhabitant (GDP/POP)						
ENV1	CO2 in tons per inhabitant (CO2/POP)						
ENV2	CO2 in kg per \$ of GDP (CO2/GDP)						

Coefficient de spécialisation

Et quand le tableau n'est pas un tableau de contingence ? La standardisation

$$X'(i) = (X(i) - \bar{X}(i))/\sigma(x)$$

Name	DEM1	DEM2	DEM3	ECO1	ECO2	ENV1	ENV2
France	0.6	-0.1	-0.4	1.2	0.9	0.4	-0.1
Italy	2.0	2.0	-0.7	0.5	0.8	0.8	0.0
Spain	0.1	-1.0	-0.7	0.0	0.2	0.3	0.1
Algeria	-1.0	0.7	1.8	-1.4	-1.5	-0.8	1.3
Libya	-1.2	-0.7	1.6	-1.3	-1.2	1.2	2.6
Morocco	-0.2	-0.1	1.0	-1.4	-1.4	-1.6	-0.2
Tunisia	-0.3	-1.1	0.8	-1.3	-1.2	-1.3	-0.1
moyenne	0	0	0	0	0	0	0
écart-type	1	1	1	1	1	1	1

TP

- ❶ Calculez le coefficient de spécialisation de chaque département en matière d'agriculteurs. Quels sont les départements où les agriculteurs sont sur-représentés ? Quels sont les départements où les agriculteurs sont sous-représentés ?
- ❷ Faites ce calcul pour chaque CSP. Quelle CSP semble présenter la plus forte disparité dans sa répartition ? Quelle CSP semble au contraire répartie de la manière la plus homogène ?
- ❸ Représentez le coefficient de spécialisation concernant les agriculteurs. Enregistrez le résultat au format AI.
- ❹ Déterminez les légendes correspondantes. A l'aide d'illustrator, finalisez cette carte.

Corrélation et régression linéaire

Les statistiques multivariées

Définition

En statistiques, les analyses multivariées ont pour caractéristique de s'intéresser à la distribution conjointe de plusieurs variables. Les analyses bivariées sont des cas particuliers à deux variables.

Les analyses multivariées sont très diverses selon l'objectif recherché ou la nature des variables. On peut identifier deux grandes familles :

- celle des méthodes descriptives visant à structurer et résumer l'information ;
- celle des méthodes explicatives visant à expliquer une ou des variables dites "dépendantes" (variables à expliquer) par un ensemble de variables dites "indépendantes" (variables explicatives).

Corrélation et régression linéaire

Définitions

Corrélation

Etudier la corrélation entre deux ou plusieurs variables, c'est étudier l'intensité de la liaison qui peut exister entre ces variables.

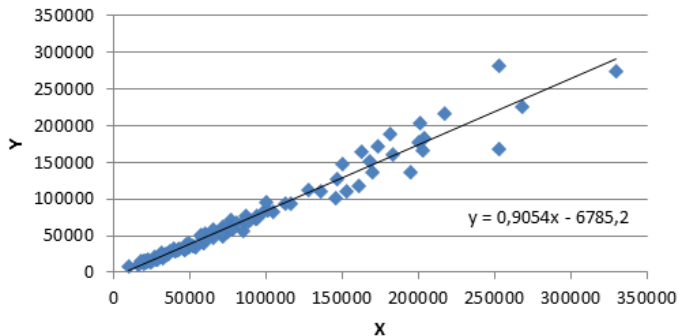
Régression linéaire

La régression est un ensemble de méthodes statistiques très utilisées pour analyser la relation d'une variable par rapport à une ou plusieurs autres.

Corrélation et régression linéaire

Des graphiques plutôt que des définitions ou des calculs pour comprendre

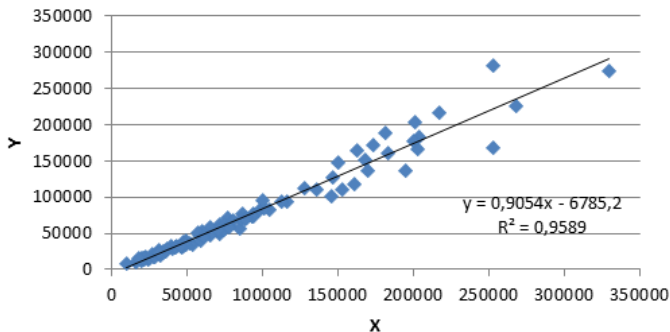
Pour faire simple, lorsque l'on étudie deux variables quantitatives, on peut produire un "nuage de points", une régression linéaire vise alors à résumer ce nuage de points par une forme plus simple à interpréter : une droite.



Corrélation et régression linéaire

Des graphiques plutôt que des définitions ou des calculs pour comprendre

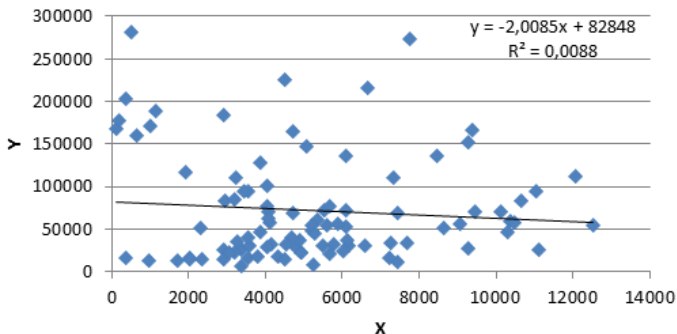
C'est le coefficient de corrélation ou le coefficient de détermination qui nous permet de dire si cette régression est "juste" :



Corrélation et régression linéaire

Des graphiques plutôt que des définitions ou des calculs pour comprendre

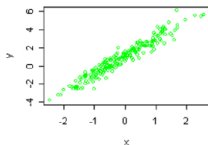
C'est le coefficient de corrélation ou le coefficient de détermination qui nous permet de dire si cette régression est "juste" ou pas du tout :



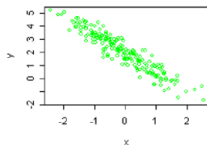
Corrélation et régression linéaire

Différentes formes de nuages de points

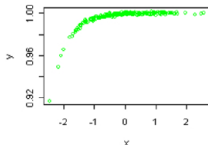
Liaison linéaire positive



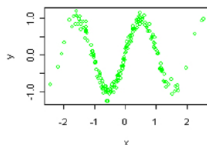
Liaison linéaire négative



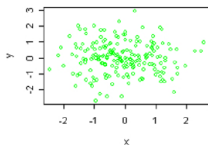
Liaison monotone positive non linéaire



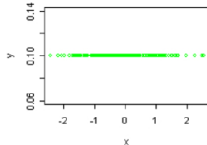
Liaison non monotone non linéaire



Absence de liaison



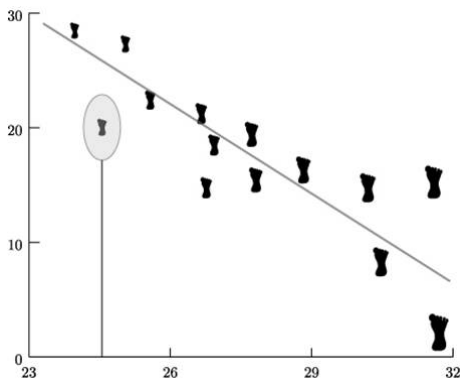
Absence de liaison



Corrélation et régression linéaire

Les pièges à éviter : Des relations problématiques

Nombre de fautes d'orthographe en fonction de la pointure. Les élèves ayant les plus grands pieds font moins de fautes.



Corrélation et régression linéaire

Les pièges à éviter : Attention à l'erreur écologique

En géographie, l'étude des corrélations se fait souvent à travers l'analyse d'un ensemble de lieux.

Lorsque les variables décrivant ces lieux sont des attributs sociaux décrivant les habitants, il faut toujours faire attention au fait qu'une corrélation établie au niveau des lieux n'implique pas forcément une corrélation au niveau des individus.

Une étude menée au niveau des individus (sociologique) peut montrer que le taux de criminalité est plus élevé chez les autochtones que chez les étrangers.

Pourtant, cette étude au niveau des quartiers (géographique) peut montrer une corrélation parfaite entre la proportion d'étrangers des quartiers et leur taux de criminalité.

Corrélation et régression linéaire

TP

- 1 Calculez la corrélation entre la variable Employé et celle de Profession Intermédiaire. Faites de même entre la variable Employé et celle d'Agriculteur. Commentez les résultats obtenus.
- 2 Calculez les résidus de la relation Employé, Profession Intermédiaire.
- 3 Représentez ces résidus sur une carte après les avoir interprétés.
- 4 Calculez la corrélation entre la variable $Tx_Employé$ et celle de $Tx_Profession\ Intermédiaire$. Commentez les résultats obtenus.

La classification ascendante hiérarchique

Présentation

La classification ascendante hiérarchique (CAH) est une méthode de classification multivariée itérative dont le principe est simple :

- On commence par calculer la dissimilarité entre les objets.
- Puis on regroupe les deux objets dont le regroupement minimise un critère d'agrégation donné (ceux qui se ressemblent le plus), créant ainsi une classe comprenant ces deux objets.
- On calcule ensuite la dissimilarité entre cette classe et les autres objets en utilisant un critère d'agrégation.
- Puis on regroupe les deux objets ou classes d'objets dont le regroupement minimise le critère d'agrégation.

On continue ainsi jusqu'à ce que tous les objets soient regroupés.

La classification ascendante hiérarchique

Prenons un exemple simple

	X_i (km)	Y_i (km)
Paris	600	2428
Marseille	846	1815
Saint-Etienne	760	2050
Bordeaux	369	1986
Reims	723	2474
Lyon	794	2087

$$Distance(euclidienne) = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$$

$$Dist_{(Paris-Marseille)} = \sqrt{(600 - 846)^2 + (2428 - 1815)^2} = 660$$

La classification ascendante hiérarchique

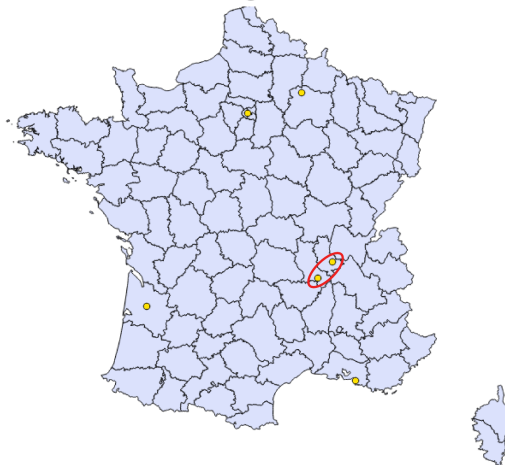
Tableau de dissimilarité (Tableau de distance)

	Paris	Marseille	Saint-Etienne	Bordeaux	Reims	Lyon
Paris	0	660	410	498	131	392
Marseille	660	0	250	506	670	276
Saint-Etienne	410	250	0	396	425	50
Bordeaux	498	506	396	0	602	436
Reims	131	670	425	602	0	393
Lyon	392	276	50	436	393	0

La classification ascendante hiérarchique

Regrouper les éléments qui sont proches

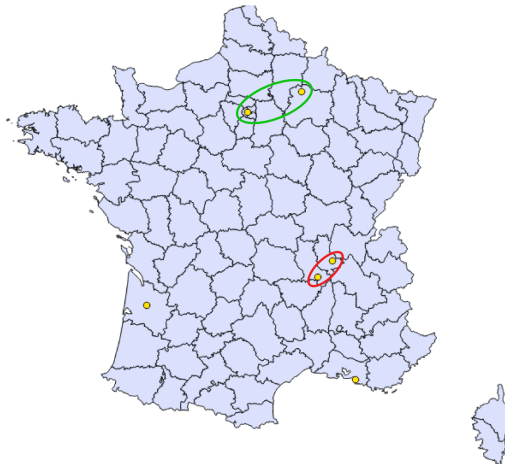
	Paris	Marseille	Saint-Etienne	Bordeaux	Reims	Lyon
Paris	0	660	410	498	131	392
Marseille	660	0	250	506	670	276
Saint-Etienne	410	250	0	396	425	50
Bordeaux	498	506	396	0	662	436
Reims	131	670	425	662	0	393
Lyon	392	276	50	436	393	0



La classification ascendante hiérarchique

Regrouper les éléments qui sont proches

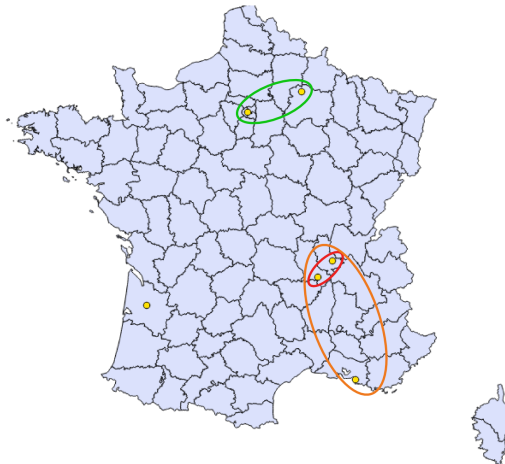
	Paris	Marseille	Saint-Etienne	Bordeaux	Reims	Lyon
Paris	0	660	410	498	<u>131</u>	392
Marseille	660	0	250	506	670	276
Saint-Etienne	410	250	0	396	425	50
Bordeaux	498	506	396	0	662	436
Reims	131	670	425	662	0	393
Lyon	392	276	50	436	393	0



La classification ascendante hiérarchique

Regrouper les éléments qui sont proches

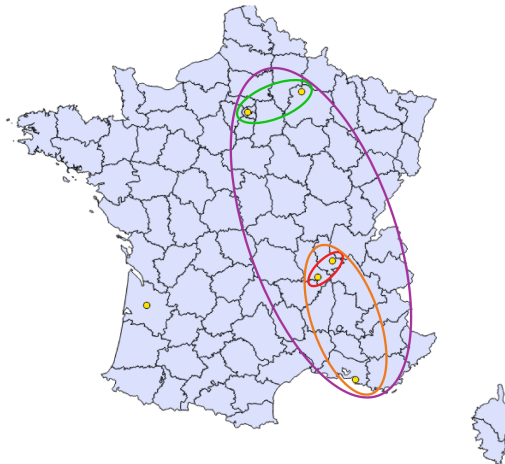
	Paris	Marseille	Saint-Etienne	Bordeaux	Reims	Lyon
Paris	0	660	410	498	131	392
Marseille	660	0	250	506	670	276
Saint-Etienne	410	250	0	396	425	50
Bordeaux	498	506	396	0	662	436
Reims	131	670	425	662	0	393
Lyon	392	276	50	436	393	0



La classification ascendante hiérarchique

Regrouper les éléments qui sont proches

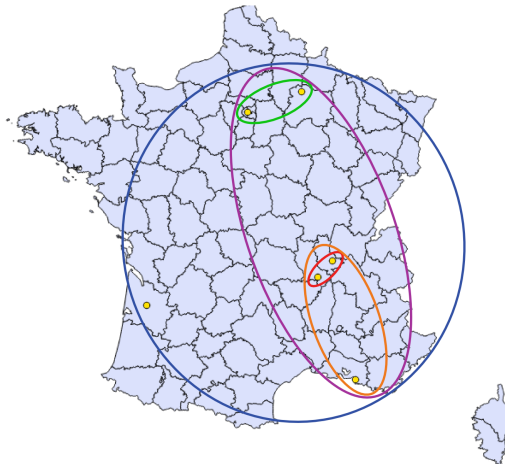
	Paris	Marseille	Saint-Etienne	Bordeaux	Reims	Lyon
Paris	0	660	410	498	131	392
Marseille	660	0	250	506	670	276
Saint-Etienne	410	250	0	396	425	50
Bordeaux	498	506	396	0	662	436
Reims	131	670	425	662	0	393
Lyon	392	276	50	436	393	0



La classification ascendante hiérarchique

Regrouper les éléments qui sont proches

	Paris	Marseille	Saint-Etienne	Bordeaux	Reims	Lyon
Paris	0	660	410	498	131	392
Marseille	660	0	250	506	670	276
Saint-Etienne	410	250	0	396	425	50
Bordeaux	498	506	396	0	662	436
Reims	131	670	425	662	0	393
Lyon	392	276	50	436	393	0



La classification ascendante hiérarchique

Le dendrogramme

Un dendrogramme est la représentation graphique d'une classification ascendante hiérarchique.

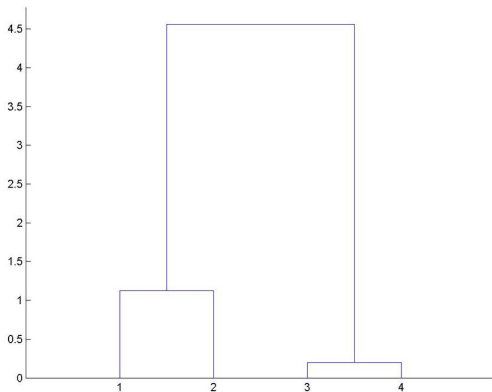
Il se présente souvent comme un arbre binaire dont les feuilles sont les individus alignés sur l'axe des abscisses.

Lorsque deux classes ou deux individus se rejoignent avec l'indice d'agrégation, des traits verticaux sont dessinés de l'abscisse des deux classes jusqu'à l'ordonnée, puis ils sont reliés par un segment horizontal.

À partir d'un indice d'agrégation, on peut tracer une droite d'ordonnée qui permet de voir une classification sur le dendrogramme.

La classification ascendante hiérarchique

Le dendrogramme : choisir un niveau de proximité pour obtenir un nombre de classes



La classification ascendante hiérarchique

Quand on a plus de deux variables c'est pas plus compliqué

	Variable 1	Variable 2	Variable 3	Variable 4
Objet 1	5	2	6	4
Objet 2	2	5	2	4

$$Dist_{(Objet1-Objet2)} = \sqrt{(5-2)^2 + (2-5)^2 + (6-2)^2 + (4-4)^2}$$

La classification ascendante hiérarchique

Plusieurs possibilités pour calculer la distance entre Paris et Marseille

Distance euclidienne : $\sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$

$$De_{(P-M)} = \sqrt{(600 - 846)^2 + (2428 - 1815)^2} = 660$$

Distance de Manhattan : $|X_1 - X_2| + |Y_1 - Y_2|$

$$Dm_{(P-M)} = |600 - 846| + |2428 - 1815| = 246 + 613 = 859$$

Distance de Tchebychev : $Max[(X_1 - X_2); (Y_1 - Y_2)]$

$$Dt_{(P-M)} = Max[(600 - 846); (2428 - 1815)] = Max[246; 613] = 613$$

La classification ascendante hiérarchique

Une précaution importante : la standardisation

Name	DEM1	DEM2	DEM3	ECO1	ECO2	ENV1	ENV2
France	107	303	12	26200	19100	6.6	0.35
Italy	192	529	9	20100	18000	7.6	0.42
Spain	78	205	9	14500	14100	6.4	0.45
Algeria	13	385	30	1500	3000	3.3	1.12
Libya	3	238	28	2000	4800	8.8	1.83
Morocco	63	294	23	1300	3600	1.1	0.29
Tunisia	58	194	22	2100	5300	1.8	0.34
West. Medit.	38	310	15	14800	13000	5.5	0.42
Definition of variables							
DEM1	Gross population density in inh/km2 (POP/SUP)						
DEM2	Net population density in inh/km2 (POP/AGR)						
DEM3	Birth rate (BIR/POP)						
ECO1	GNP in \$ per inhabitant (GNP/POP)						
ECO2	GDP in p.p.a per inhabitant (GDP/POP)						
ENV1	CO2 in tons per inhabitant (CO2/POP)						
ENV2	CO2 in kg per \$ of GDP (CO2/GDP)						

La classification ascendante hiérarchique

Une précaution importante : la standardisation

Name	DEM1	DEM2	DEM3	ECO1	ECO2	ENV1	ENV2
France	0.6	-0.1	-0.4	1.2	0.9	0.4	-0.1
Italy	2.0	2.0	-0.7	0.5	0.8	0.8	0.0
Spain	0.1	-1.0	-0.7	0.0	0.2	0.3	0.1
Algeria	-1.0	0.7	1.8	-1.4	-1.5	-0.8	1.3
Libya	-1.2	-0.7	1.6	-1.3	-1.2	1.2	2.6
Morocco	-0.2	-0.1	1.0	-1.4	-1.4	-1.6	-0.2
Tunisia	-0.3	-1.1	0.8	-1.3	-1.2	-1.3	-0.1
moyenne	0	0	0	0	0	0	0
écart-type	1	1	1	1	1	1	1

La classification ascendante hiérarchique

Une précaution importante : la standardisation

Données

Name	DEM1	DEM3
France	0.6	-0.4
Italy	2.0	-0.7
Spain	0.1	-0.7
Algeria	-1.0	1.8
Libya	-1.2	1.6
Morocco	-0.2	1.0
Tunisia	-0.3	0.8



Distance euclidienne sur variables normées

	Fra	Ita	Spa	Alg	Lib	Mor	Tun
Fra	0.0	1.5	0.6	2.7	2.6	1.5	1.5
Ita	1.5	0.0	1.9	4.0	4.0	2.8	2.8
Spa	0.6	1.9	0.0	2.8	2.6	1.7	1.6
Alg	2.7	4.0	2.8	0.0	0.3	1.2	1.2
Lib	2.6	4.0	2.6	0.3	0.0	1.2	1.2
Mor	1.5	2.8	1.7	1.2	1.2	0.0	0.1
Tun	1.5	2.8	1.6	1.2	1.2	0.1	0.0

TP

- 1 A l'aide de Philcarto, produisez une CAH sur les CSP des départements français.
- 2 Commencez par interpréter les résultats obtenus pour une partition en deux classes. Allez comme cela jusqu'à une partition en six classes. Enregistrez la carte au format ai.
- 3 Finalisez la carte sous Illustrator.